at beginning of transition, a is specified in said bookmark, g_s is a specified parameter, and t is time, ranging between 0 and 1.

70. The method of claim 63 where the FAP amplitude transition path follows the equation

$FAPAmp(t) = startVal_i(2t^3 - 3t^2 + 1) + FAPval_i(-2t^3 + 3t^2) + startTan_i(t^3 - 2t^2 + 1)$, where *startVal*, *FAPval*, and *startTan*, are specified constants.

--.

REMARKS

The above-identified Office action consists of a "cut-and-paste" copy of the rejections in the last Office action, except for section 8, which is a "Response to arguments."

In applicants' previous remarks, the teachings of the references were described and distinguished from the claims. Those remarks are hereby incorporated by reference.

The Examiner's "response to arguments" states that applicants' remarks are not persuasive because

1) with respect to the combination of audio/video signals, examiner argues that the rendering system of Goldenthal is the apparatus that combines the signal;
2) with respect to the claimed FAP information, applicant does not compare/contrast this information versus Goldenthal's facial information/instructions;
3) Goldenthal teaches text via the 'chat room' application – internet chat rooms inherently contain text

This is the sole explanation for holding that applicants' argument are not persuasive and, applicants respectfully submit, this explanation does not justify the holding.

Clearly, the "response to arguments" addresses **only** the Goldenthal reference. Therefore, since the "response to arguments" is completely silent as to applicants' arguments about the Gasper reference, and Chen et al reference, applicants respectfully submit that either *all claims that are rejected in view of Gasper by itself, or in view of a combination that includes the Chen reference should be allowed, or a justifiable reason for holding that applicants' arguments are not persuasive is called for.*

In connection with the Goldenthal reference, and particularly in connection with the rejection of claim 2 as being anticipated by Goldenthal, in the last-filed amendement applicants asked the question: "what does the Examiner assert to be the 'apparatus' in Goldenthal et al that corresponds to the 'apparatus' of claim 2?" Since claim 2 defines a method for transmitting signals to an apparatus, and that means that the signal is created elsewhere and is transmitted to the apparatus, applicants felt it is essential to know what the Examiner considers the apparatus TO which the method transmits a signal.

Item 1 in the Examiner's "response to arguments" appears to be directed to that applicants' question. The Examiner's states that "the rendering system of Goldenthal is the apparatus that combines the signal[s]." Unfortunately, since claim 2 specifies an apparatus, and also specifies a step of combining signals, the answer to applicants' question is left unclear because there are two different interpretations to the Examiner's statement:

(1) the Examiner argues that the rendering system of Goldenthal is the apparatus that combines the signals and which thus creates the signal that is transmitted to the apparatus mentioned in claim 2, or

(2) the Examiner argues that the rendering system of Goldenthal is the apparatus referred to in claim 2, i.e., the apparatus to which the created signal is being transmitted.

The question is: which is the correct interpretation of the Examiner's remark?

Since rendering system 240 does have an input from line 116 that supplies the visemes, and an input arriving from audio file 117, the first interpretation appears the more plausible one. Therefore, the following remarks assume that element 240 is the apparatus where the signal corresponding to the step of combining **is created**, making the apparatus **to which** the combined signal is transmitted be composed of display screen 2 and loudspeaker 3.

Even with this interpretation, however, applicants respectfully submit that claim 2 is not anticipated or rendered obvious by Goldenthal.

1. Elements 2 and 3 cannot be the apparatus referred to in claim 2 because the "signal stream for said transmitting" specified in claim 2 is not transmitted to these elements. Rather, each of the elements receives its own signal.

2. Elements 2 and 3 cannot be the apparatus referred to in claim 2 because the term "transmitting signals" is *not* employed in the art to the hardware connection between an output circuit and its associated speaker and screen display
3. There is no teaching whatsoever that the signal of line 116 and the signal arriving from audio file 117 are **combined** within rendering system 240 (or anywhere else, for that matter) to form the "signal stream for said transmitting" defined in claim 2. It is noted that, in fact, there is no teaching at all of how rendering system 240 can be realized, other than a reference to US Patent 5,657,426, issued to Waters et al. That patent, however, is responsive solely to a text string and does not teach creating a first signal that is like the first signal of claim 1, or creating a second signal like the second signal of claim 2. Consequently, it is not surprising that it does not teach creating a combined signal as defined in claim 2. Actually, that patent does not teach creating any signal that combines audio and video information, since what it does describe is two distinct paths: one for audio, and the other for video.

It is respectfully submitted, therefore, that claim 2 is not anticipated or rendered obvious by Goldenthal.

Applicants respectfully submit that the language of claim 2, which speaks in terms of transmitting a signal – based on the meaning commonly, and consistently, used in the art to designate the communication of signals a significant distance over a communication network, and NOT to designate the movement of signals within a circuit or an apparatus – applies much more closely to Goldenthal's FIG. 3, than to FIG. 1. In FIG. 3, at least, the step of transmitting a signal to apparatus is present. Of course, the system described by Goldenthal in FIG. 3 fails to meet the other limitations of claim 2; in particular, the ultimate and the penultimate steps of the claim. Therefore, it is respectfully submitted that claim 2 is neither anticipated nor rendered obvious by Goldenthal et al.

As for claim 12, which is the next independent claim (rejected over Goldenthal), it defines an apparatus. If the rendering system of Goldenthal -- i.e., element 240 -- corresponds to the apparatus of claim 12 (as the Examiner appears to argue), then to sustain the rejection this element 240 must contain all of the limitations that claim 12 specifies. That, however, is not the case.

Claim 12 specifies:

> a decoder, responsive to an input signal comprising signals representing audio and embedded video synthesis command signals, that separates the command signals from signals representing audio to develop an audio signal stream and a video synthesis command signals stream,

but no such decoder is described for rendering system 240. There is no description of a signal representing audio and embedded video synthesis command signals (and the Examiner has not pointed to any) and there is no description of any decoder (and the Examiner has not pointed to any).

Claim 12 also specifies

> a converter responsive to said audio signal stream for developing sound, and
>
> a video synthesizer responsive to said video synthesis command signals stream for developing images.

Here, too, the reference provides no description of these elements, although the reference does refer to the above-mentioned Waters et al patent. The Waters et al patent has elements that correspond to a converter and to a video synthesizes, but the converter element in the Waters et al patent is NOT responsive to the audio signal stream portion of a signal that comprises signals representing audio **and** embedded video synthesis command signals, and the video synthesizer element in the Waters et al patent is NOT responsive to the video synthesis command signals stream portion of a signal that comprises signals representing audio **and** embedded video synthesis command signals.

It is respectfully submitted, therefore, that claim 12 is neither anticipated nor rendered obvious by Goldenthal. Consequently, claims 26-28, which depend on claim 12, are also neither anticipated nor rendered obvious by Goldenthal.

The next independent claim (rejected over Goldenthal) is claim 31. It specifies a method that receives a certain input signal, decomposes it into two streams, and processes each of the streams. While applicants admit that the reference *suggests* that system 240 processes an audio signal stream (the stream for element 117) and also processes a video stream (arriving on line 116) -- since such signals are applied to element 240 -- it is respectfully submitted that it does not receive the signal (in singular) specified in claim 31, and it does not perform the step of separating defined in claim 31. Therefore, it is

respectfully submitted that claim 31 is neither anticipated nor rendered obvious by Goldenthal. Claims 32, 34, and 36-40 depend on claim 31.

As an aside, it is noted that claim 12 defines a method that is carried out effectively in an encoder, and claim 31 defines a method that is carried out effectively in a decoder. Since an encoder and a decoder do not perform the same function, it is not possible for element 240 to be both. The Examiner is constrained to assert one, or the other, or neither – but not both.

In light of the above, applicants respectfully submit that:

- The Examiner failed not only to rebut applicants argument relative to the Gasper reference and relative to the Chen reference, but has even failed to give <u>any</u> reason for holding that applicants' arguments were not persuasive.
- The Examiner's "response to arguments" relative to the Goldenthal reference, which contained 3 items, is lacking.
    - The first item answers a question, and that does not constitute a rebuttal of applicants' argument, or a reason for holding an argument not persuasive.
    - The second item impliedly finds fault with applicants' argument, but that which the Examiner finds lacking in applicants' argument would have been irrelevant to the argument. Therefore, the second item does not rebut applicants' arguments and constitutes an unjustifiable reason to hold that applicants' argument is unpersuasive.
    - Lastly, the third item assets that Goldenthal teaches something, but applicants believe that this teaching is wholly irrelevant to the rejected claims. Therefore, the third item also does not rebut applicants' arguments and constitutes an unjustifiable reason to hold that applicants' argument is unpersuasive.

Claims 1, and 12-25 were rejected under 35 USC 102 as being anticipated by Gasper. As indicated above, the Examiner has not rebutted applicants' previous arguments regarding Gasper and, therefore, applicants are unable to determine in what aspect of applicants' argument the Examiner found the arguments not persuasive. Applicant is looking forward to an explanation from the Examiner in the next Office Action.

Nevertheless, it might be helpful to address the issue afresh.

Gasper does not describe a system that deals with a model of a talking head, and the image that is created is not developed through application of parameters to a model. Further there is no input signal that comprises text and FAP information, and there is no decoder that separates the FAP information from the text and also develops phonemes from said text. The Examiner cites element 10 in FIG. 1 of Gasper. This element is a microprocessor, but there is no teaching anywhere in Gasper that (a) one input to microprocessor 10 is an "input signal comprising text and FAP information," and that (b) microprocessor 10 "separates the FAP information from the text, and develops phonemes from said text." In short, the decoder defined in claim 1 is absent in the Gasper reference. Therefore, claim 1 is neither anticipated nor made obvious by the Gasper reference.

In fact, element 10 of Gasper does not even receive text. Rather, the tiled letters (qua "input text") are converted in text-to-phoneme translator 40 to a phoneme string, prior to any action by element 10. As indicated previously, the Examiner would be more correct to equate the decoder of claim 1 with translator 40 of Gasper rather than with element 10 Gasper, although that would still not correspond to the decoder of claim 1 because of the absence of FAP information, and any treatment of FAP information in translator 40 (or any other information other than the input text).

Additionally, claim 1 specifies a converter that converts the phonemes to additional FAP information and outputs the additional FAP information together with the FAP information separated by the decoder. No such converter exists in the Gasper system. The Examiner pointed to element 26, this element is an audio generator, and the converter element of claim 1 does not even deal with audio matters.

Additionally still, claim 1 specifies a face rendering module that is responsive to an applied face model signal and to the output of the converter. The Examiner cited element 18, which is a video generator. As indicated previously, element 18 is simply the conventional module of a computer that creates the signals that drive a monitor, and it has no algorithmic processing .

In view of the above analysis, it is respectfully submitted that **none** of the element defined by claim 1 elements is anticipated, or made obvious, by the teachings of Gasper.

Applying the above arguments to claims 12-25, applicants respectfully submit that claims 12-25 are also not anticipate or rendered obvious by Gasper.

As for Chen et al, they focus their description of reducing the amount of data that is contained in a video signal, and through analysis they reduce the amount of data. This data is not synchronized to voice because it is created solely from a video signal. In light of this, combining the teachings of Chen et al and the teachings of Gasper not yield the system defined in claims 24 and 25; especially given the very explicit definitions found in claims 24 and 25.

Likewise, combining the teachings of Chen et al and the teachings of Goldenthal doe not yield the system defined in claims 29, 30, 41 and 42, both because of the deficiencies in Goldenthal, as discussed above, and because of the specific limitations found in claims 29, 30, 41 and 42.

New claims 43-70 are introduced. Based on the above analysis of the cited references, applicants believe that the claims are neither anticipated nor made obvious by the prior art.

A Declaration under 37 CFR 1.132 is respectfully submitted herewith by Dr. Hans Peter Graf – who is a co-inventor inventor of US 6,028,960 that was cited by the Examiner. It is respectfully submitted that the Declaration expresses Dr. Graf's views about the teachings contained in the Goldenthal, Gasper, and Chen et al references, and those teachings match perfectly applicants' assertions contained in this, and previous, responses, traversing the Examiner's rejections.

In light of the above, it is respectfully submitted that all of the Examiner's rejections have been overcome. Reconsideration and allowance are earnestly solicited.

Respectfully,
Mark Beutnagel
Ariel Fischer
Joern Ostermann
Yao Wang

Dated: 10/14/02

By _____
Henry T. Brendzel
Reg. No. 26,844
Phone (973) 467-2025

Appendix Marked up version, showing changes made

In the claims:

Please add the following claims: --

43. Apparatus comprising

A decoder/synthesize module that is responsive to an input that includes a text specification and explicit FAP information, outputting a synthesized voice at a first output, and phonemes as well as said FAP information at a second output;

a converter responsive to said second output for generating a sequence of facial animation parameters;

face rendering module responsive to said converter; and

a compositor, responsive to said synthesizer and to said face rendering module.

44. The apparatus of claim 43, further adapted to accept said input from a remote location that is communicated to said apparatus via a communication network.

45. The apparatus of claim 43 where said FAP information that is explicitly included in said input comprises interspersed bookmarks.

46. The apparatus of claim 45 where each bookmark conveys information about identity of a FAP, and ultimate state of the FAP.

47. The apparatus of claim 46 where said information conveys amplitude information

48. The apparatus of claim 46 where said information conveys a duration measure for transiting to specified state.

49. The apparatus of claim 46 where the said ultimate state of the FAP is reached in accordance with a specified transition path.

50. The apparatus of claim 49 where the transition path is selected by said facial animation module.

51. The apparatus of claim 49 where said transition path is specified by the bookmark.

52. The apparatus of claim 49 where the transition path follows the equation $f(t) = a_s + (a - a_s)t$, where a_s is amplitude measure at beginning of transition, a is specified in said bookmark, and t is time, ranging between 0 and 1.

53. The apparatus of claim 49 where the transition path follows the equation $f(t) = a_s + (1 - e^{-t})(a - a_s)$, where a_s is amplitude measure at beginning of transition, a is specified in said bookmark, and t is time, ranging between 0 and 1.

54. The apparatus of claim 49 where the transition path follows the equation $f(t) = a_s + \frac{(a - a_s)}{(1 - e^{-\lambda(t - FAPdur/2)})}$, where a_s is amplitude measure at beginning of transition, a is specified in said bookmark, *FABdur* is specified in said bookmark, λ is a specified parameter, and t is time, ranging between 0 and 1.

55. The apparatus of claim 49 where the transition path follows the equation $f(t) = a_s + (2t^3 - 3t^2 + 1) + (-2t^3 + 3t^2)a + (t^3 - 2t^2 + t)g_s$, where a_s is amplitude measure at beginning of transition, a is specified in said bookmark, g_s is a specified parameter, and t is time, ranging between 0 and 1.

56. The apparatus of claim 49 where the FAP amplitude transition path follows the equation $FAPAmp(t) = \underset{i}{startVal}(2t^3 - 3t^2 + 1) + \underset{i}{FAPval}(-2t^3 + 3t^2) + \underset{i}{startTan}(t^3 - 2t^2 + 1)$, where *startVal*, *FAPval*, and *startTan*, are specified constants.

57. A method comprising the steps of:

receiving an input that includes a text specification and explicit FAP information, and outputting a synthesized voice at a first output, and phonemes as well as said FAP information at a second output;

generating a sequence of facial animation parameters from signals of said second output;

rendering images from output signals developed by said step of generating; and

a combining said synthesized voice and said images.

58. The method of claim 57, where said step of receiving accepts said input from a remote location that is communicated to said apparatus via a communication network.

59. The method of claim 57 where said FAP information that is explicitly included in said input comprises interspersed bookmarks.

60. The method of claim 59 where each bookmark conveys information about identity of a FAP, and ultimate state of the FAP.

61. The method of claim 60 where said information conveys amplitude information

62. The method of claim 60 where said information conveys a duration measure for transiting to specified state.

63. The method of claim 60 where the said ultimate state of the FAP is reached in accordance with a specified transition path.

64. The method of claim 63 where the transition path is selected by said facial animation module.

65. The method of claim 63 where said transition path is specified by the bookmark.

66. The method of claim 63 where the transition path follows the equation $f(t) = a_s + (a - a_s)t$, where a_s is amplitude measure at beginning of transition, a is specified in said bookmark, and t is time, ranging between 0 and 1.

67. The method of claim 63 where the transition path follows the equation $f(t) = a_s + (1 - e^{-t})(a - a_s)$, where a_s is amplitude measure at beginning of transition, a is specified in said bookmark, and t is time, ranging between 0 and 1.

68. The method of claim 63 where the transition path follows the equation $f(t) = a_s + \frac{(a - a_s)}{(1 - e^{-\lambda(t - FAPdur/2)})}$, where a_s is amplitude measure at beginning of transition, a is specified in said bookmark, *FABdur* is specified in said bookmark, λ is a specified parameter, and t is time, ranging between 0 and 1.

69. The method of claim 49 where the transition path follows the equation $f(t) = a_s + (2t^3 - 3t^2 + 1) + (-2t^3 + 3t^2)a + (t^3 - 2t^2 + t)g_s$, where a_s is amplitude measure at beginning of transition, a is specified in said bookmark, g_s is a specified parameter, and t is time, ranging between 0 and 1.

70. The method of claim 63 where the FAP amplitude transition path follows the equation $FAPAmp(t) = \underset{i}{startVal}(2t^3 - 3t^2 + 1) + \underset{i}{FAPval}(-2t^3 + 3t^2) + \underset{i}{startTan}(t^3 - 2t^2 + 1)$, where *startVal*, *FAPval*, and *startTan*, are specified constants.

--.